

IOWA STATE UNIVERSITY

Digital Repository

World Languages and Cultures Publications

World Languages and Cultures


2018

Examining the temporal structure of the perception-production link in SLA: A longitudinal study

Charles Nagle

Iowa State University, cnagle@iastate.edu

Follow this and additional works at: http://lib.dr.iastate.edu/language_pubs

 Part of the [Educational Assessment, Evaluation, and Research Commons](#), [Educational Methods Commons](#), and the [Language and Literacy Education Commons](#)

The complete bibliographic information for this item can be found at http://lib.dr.iastate.edu/language_pubs/133. For information on how to cite this item, please visit <http://lib.dr.iastate.edu/howtocite.html>.

This Article is brought to you for free and open access by the World Languages and Cultures at Iowa State University Digital Repository. It has been accepted for inclusion in World Languages and Cultures Publications by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.

Examining the temporal structure of the perception-production link in SLA: A longitudinal study

Abstract

Most studies on the perception-production link have assumed a synchronous relationship, according to which gains in perception transfer to production rapidly and efficiently. However, time-lagged and asymptotic relationships are also possible, in which case perception would guide production at a later stage or production would improve only once perception has reached a high level of accuracy. The present study investigates the temporal dynamics of the perception-production link by modeling English speakers' ability to perceive and produce L2 Spanish stops over time. Mixed-effects modeling of the d' perception and voice onset time (VOT) production data demonstrated significant development in both areas. Time-lagged change models indicate that a change in d' was significantly related to decreasing VOT in L2 /p/ at the following session, but no significant relationships emerged between perception and production of L2 /b/. A range of patterns emerged at the individual level. Results are interpreted in light of contemporary models of L2 speech learning.

Keywords

Pronunciation, longitudinal, mixed-effects modeling, speech perception, speech production, Spanish

Disciplines

Educational Assessment, Evaluation, and Research | Educational Methods | Language and Literacy Education

Comments

This article has been accepted for publication in *Language Learning* (2017). Posted with permission.

Examining the temporal structure of the perception-production link in SLA: A longitudinal study

Charles L. Nagle

Iowa State University

*This article has been accepted for publication in *Language Learning*

Author Note

I would like to express my deepest gratitude to Pavel Trofimovich for his encouragement and insightful feedback on earlier versions of this paper. I would also like to thank the anonymous reviewers for their invaluable feedback.

Correspondence concerning this article should be addressed to Charles Nagle, World Languages and Cultures, 3102G Pearson Hall, 505 Morrill Road, Ames, Iowa 50011.

Email: cnagle@iastate.edu.

Abstract

Most studies on the perception-production link have assumed a synchronous relationship, according to which gains in perception transfer to production rapidly and efficiently. However, time-lagged and asymptotic relationships are also possible, in which case perception would guide production at a later stage or production would improve only once perception has reached a high level of accuracy. The present study investigates the temporal dynamics of the perception-production link by modeling English speakers' ability to perceive and produce L2 Spanish stops over time. Mixed-effects modeling of the d' perception and voice onset time (VOT) production data demonstrated significant development in both areas. Time-lagged change models indicate that a change in d' was significantly related to decreasing VOT in L2 /p/ at the following session, but no significant relationships emerged between perception and production of L2 /b/. A range of patterns emerged at the individual level. Results are interpreted in light of contemporary models of L2 speech learning.

Keywords: Pronunciation; longitudinal; mixed-effects modeling; speech perception; speech production; Spanish.

Introduction

During the first year of life, speech perception becomes attuned to ambient language, such that between 6 and 12 months infants discriminate native language (L1) contrasts with increasing accuracy while their ability to discriminate non-native contrasts diminishes (Kuhl et al., 2006). By the end of the first year, native contrasts are perceived as more distinct than non-native pairs irrespective of the absolute acoustic distance between constituents, which indicates that phonological categories have begun to coalesce (Cheour et al., 1998). Kuhl and colleagues have demonstrated that perceptual attunement to the native language fundamentally restructures phonetic space through perceptual warping: L1 prototypes begin to act as perceptual magnets, drawing in adjacent acoustic tokens (Kuhl, 1991, 2000; Kuhl et al., 2008; Kuhl, Conboy, Padden, Nelson, & Pruitt, 2005). Perceptual warping is advantageous because it allows the listener to process the natural variation present in speech with greater efficiency by associating a wider range of acoustic targets with L1 categories. However, L1 perceptual attunement has consequences for L2 speech learning, which may be challenging if the L2 makes use of a portion of phonetic space that is near an L1 category or otherwise involves acoustic cues that are not employed in the L1. Contemporary models of L2 phonetic learning such as the Speech Learning Model (SLM; Flege, 1995) thus argue that the acquisition of L2 sounds is not constrained by a neurobiological critical period, but rather by the developmental state of the L1 system, which is typically robust by the time L2 learning begins. According to L2 models, perceived relationships between L1 and L2 phones shape L2 speech production; if L1 and L2 phones are similar, then learners may not detect the phonetic differences between them, leading to an accented L2 production.

Although many studies on the perception-production link in L2 learning have upheld this view, weak or statistically insignificant findings are not uncommon, leading some scholars to suggest that the two modalities may dissociate in the initial stages of SLA (Hanulíková, Dediu, Fang, Bašnaková, & Huettig, 2012). However, a synchronous relationship, wherein gains in perception rapidly and efficiently transfer to production, has typically been assumed, even though other temporal models are possible. For example, gains in perception may predict delayed gains in production, in which case a time-lagged model of the perception-production link would be appropriate. On the other hand, if production improves substantially only once perception has reached a high level of accuracy, then perception and production would be asymptotically related. The present longitudinal study investigates the temporal properties of the perception-production link in instructed SLA by evaluating contemporaneous, time-lagged, and asymptotic approaches to the perception and production of L2 Spanish stops. Mixed-effects models were fit to the perception and production data to examine development in each domain before exploring how changes in the former were related to changes in the latter. Individual data were also taken into consideration, complementing the group-level models.

Background

Models of L2 Speech Perception & Production

Contemporary models of L2 phonetic learning argue that perceived relationships between L1 and L2 sounds shape L2 speech perception and production. According to the SLM (Flege, 1995, 2003), speakers must detect the phonetic differences that distinguish similar sounds. If they succeed, then they will likely establish a new phonetic category, enabling them to produce nativelike L2 phones. If, on the other hand, the listener equates the two sounds, then category formation will be blocked, and the L1 category will be recycled in the L2, resulting in an

accented production. In that case, L1 and L2 phones may converge, leading to a compromise production (Williams, 1980) whose properties reflect the composition of the merged category.

Concentrating on the development of accurate L2 perceptual routines, PAM-L2 (Best & Tyler, 2007) explicates how L2 contrasts are assimilated to L1 categories. If both members of the L2 contrast are perceived as equally good exemplars of a single L1 category (a *single category* assimilation), then discrimination is predicted to be poor. Conversely, if one L2 sound is a poorer fit than the other (a *category goodness* scenario), then the L2 contrast may be easier to perceive. Although the theoretical tenets of the SLM and PAM-L2 broadly align, the models diverge in a number of important ways. Whereas the SLM is predicated on an auditory/acoustic approach to speech perception, PAM-L2 draws upon the direct realist tradition and therefore contends that listeners perceive articulatory gestures. Consequently, though PAM-L2 does not explicitly address L2 production, it could be argued that perception and production should be more tightly synchronized on the basis of its gestural approach. On the other hand, within the framework of the SLM, the speaker would need to decompose acoustic/auditory cues into their respective gestural constellations, which could result in a lagged relationship.

The L2 Linguistic Perception model (L2LP; Boersma & Escudero, 2008; Escudero, 2007; van Leussen & Escudero, 2015; Vasiliev & Escudero, 2014) assumes that individuals create a copy of the L1 perceptual system, which serves as the initial state of L2 learning. Therefore, whereas the SLM posits bidirectional cross-linguistic influence through the mechanisms of phonetic category assimilation and dissimilation, according to the L2LP, while the L2 perceptual grammar changes over time, the L1 grammar remains stable. Diverging from the other two models, the L2LP furthermore adopts a generative approach (stochastic optimality theory) to L2 perceptual learning. In this framework, constraints are defined as a continuous series of auditory

features whose organization depends on the specific characteristics of the target language.

Learning is probabilistic and meaning-driven such that learners strive to minimize perceptual errors by maximizing the likelihood that the form they have perceived coincides with the form the speaker intended. If the perceived target is not the intended target, then the connections that resulted in the perceptual error are weakened, and if perceived and intended targets coincide, then the corresponding perceptual pathways are strengthened. Through the progressive weakening and strengthening of the linkages between auditory, phonetic, phonemic, and lexical tiers, learners adjust their L2 perceptual grammar, leading to nativelike perception. Perceptual assimilation to native categories also plays a prominent role in this model. Cases in which the L2 contrast is assimilated to a single native category (a *new* scenario), requiring L1 category split or the creation of a new L2 category altogether, are predicted to be more challenging than cases in which the L2 contrast is mapped to two native categories (a *similar* scenario), which entails shifting the L2 perceptual boundary to fit L2 phonetic characteristics. Since the initial state of L2 learning is a copy of the L1 perceptual grammar, each individual begins with a unique L2 grammar derived from the specific characteristics of the L1 dialect. This may lead to individual differences in pathways to accurate L2 perception. For example, Chládková and Podlipský (2011) found that Moravian Czech speakers assimilated the Dutch /i-ɪ/ contrast to a single category, Czech /ɪ/, but Bohemian speakers exhibited a two category assimilation to Czech /i:/ and /ɪ/. These assimilations set the stage for L2 learning: Whereas Moravian speakers would need to split the L1 category or create a new L2 category, for Bohemian speakers a boundary shift would be necessary (see also, Mayr & Escudero, 2010). Though the L2LP deals exclusively with perception, production studies likewise indicate that individual differences in L1 speech patterns contribute to L2 learning. For example, Kartushina and Frauenfelder (2014) found that

L1 Spanish speakers' vowel production, specifically their L1 vowel compactness, was associated with more accurate production of L2 French vowels.

Summarizing these theoretical positions, all three models posit that cross-linguistic relationships between the L1 and L2 determine learning trajectories in perception (PAM-L2 and L2LP) and production (SLM). The tenets of the SLM suggest an asynchronous perception-production relationship. It could be that production improves as L2 phonetic categories develop, in which case the amount of delay between the two modalities could be tied to linguistic factors such as the degree of L1-L2 overlap and the articulatory complexity of the L2 target as well as individual differences in perceptual skill or aptitude. At the same time, nativelike performance would be expected only once L2 categories have become robust, though nativelike perceptual ability may not be sufficient in and of itself to facilitate nativelike production. The direct realist approach of PAM-L2 seems to indicate that perception and production could be more tightly aligned, though as with the SLM, even if learners are capable of perceiving (i.e., discriminating) the gestures involved in L2 sounds, they may struggle to enact the necessary motor commands to reproduce those gestures in speech production. Individual differences in the perception-production link could also be observed such that for some individuals, accurate perception may feed accurate production quickly and efficiently whereas for others the two modalities may appear to be dissociated.

Research on the Perception-Production Link

Scholarship has investigated the perception-production link through various experimental designs, including studies evaluating whether training in one modality produces gains in the other and cross-sectional and longitudinal studies whose goal is to explore relationships between the two modalities at various proficiency levels over time. Following theoretical claims that

accurate perception facilitates accurate production, a large body of work has examined if training the former leads to gains in the latter. For example, in Bradlow and colleagues' (1997) study, 11 L1 Japanese speakers of L2 English completed a minimal pair identification task and a repetition task to assess their perception and production of the English /r/-/l/ contrast before and after participating in high variability perceptual training. The trained group significantly outperformed a matched control group on the posttest perceptual measure, and L1 English listeners discriminated trained speakers' posttest productions significantly better than their pretest productions, which suggests that perceptual training had boosted production. In a more recent study, Hanulíková and colleagues (2012) trained L1 Dutch speakers on L2 Slovak consonant clusters. Participants completed three perception measures, phoneme monitoring, word monitoring, and mispronunciation detection, and two production tasks, imitation and reading, on three occasions over a three-week period. Underscoring the fact that perception tasks may tap into distinct abilities, the authors found that the three perception measures were only significantly correlated at the third session. Moreover, despite individual differences in performance on the perception and production measures over time, the only statistically significant correlation to emerge was between the mispronunciation detection task and reading task at the third session, which the authors interpreted as evidence of a possible disconnection or dissociation between the two modalities: "It is likely that in the initial stages of L2 acquisition, perception and production dissociate, and learners approach the task of learning these sequences with differing strategies. It could be that only once precise perceptual representations have been established, accurate utterances can be observed" (p. 99).

In addition to perception training, scholars have investigated production training and combined training paradigms. Regarding the former, Kartushina, Hervais-Adelman,

Frauenfelder, and Golestani (2015) trained L1 French speakers on four Danish vowels over five sessions using trial-by-trial visual feedback on tongue height and backness. Production was assessed via a shadowing task and perception through ABX discrimination. Significant gains in production were observed for the experimental group but not for the control group, and the experimental group's perception improved significantly, albeit modestly, at posttest. Perception and production were not correlated at pretest, but there was a nearly significant correlation between changes in perception and changes in production once an outlier was removed.

However, in some cases, production was poor despite accurate perception, leading the authors to argue that “accurate L2 speech perception is not sufficient for accurate L2 speech production” (p. 828). In fact, there is evidence that producing speech during training can actually disrupt perceptual learning. In Baese-Berk and Samuel's (2016) study, L1 Spanish speakers were assigned to perception only and perception and production training conditions targeting Basque fricative-affricate contrasts, and perceptual learning was evaluated using a 19-step acoustic continuum ranging from /sa/ to /ʃa/. The perception only group demonstrated categorical perception of the affricate contrast; within-category discrimination was poor and a peak was observed at the category boundary near the middle of the continuum. In contrast, a weak boundary emerged when the perception and production group read aloud a random letter, and no peak emerged when they were repeated one of the target stimuli, which suggests that the particular forms learners had to produce mediated the extent to which speech production disrupted perceptual learning.

Taken together, these studies demonstrate that whereas accurate production can be facilitated by training perception, activating production during training may compromise emerging perceptual representations. However, although perception training has been associated

with medium gains in perception and small but reliable gains in production, the extent to which training perception transfers to production depends on a range of factors, including the content and duration of the training paradigm, the target structure, and the tasks used to evaluate performance in each modality (Sakai & Moorman, 2017). Moreover, the fact that most training studies occur over a short period and therefore do not evaluate longer-term learning outcomes may also explain the diversity of findings in the area since gains in one modality may transfer to the other at a later stage.

Complementing laboratory training studies, a large body of work has examined the perception-production link in naturalistic language learners who did not receive targeted training (Borden, Gerber, & Milsark, 1983; Flege, Bohn, & Jang, 1997; Flege, MacKay, & Meador, 1999; Rochet, 1995). In general, this work has shown that perception and production are correlated and that they correlate more strongly in more proficient L2 users, which suggests that they become synchronized over time. For example, Flege, Bohn, and Jang (1997) found that experienced L2 English speakers made greater use of spectral cues than inexperienced speakers when perceiving and producing the English /i/-/ɪ/ and /ɛ/-/æ/ vowel contrasts. Regression analyses further revealed that perception and production were related in a contrast-specific manner insofar as perception scores for /i/-/ɪ/ accounted for more variance in the production of /i/-/ɪ/ than /ɛ/-/æ/. On the other hand, in a study on Korean speaker's acquisition of L2 English stops, Schertz, Cho, Lotto, and Warner (2015) found that cue use in production did not reflect cue weighting in perception. In particular, whereas participants mostly relied on f_0 at vowel onset when labeling stops, VOT was a more reliable predictor of stop consonant identity in production. Thus, although participants successfully discriminated the stops in both perception and production, at a more granular phonetic level, the two modalities were not aligned.

A smaller set of longitudinal studies has tracked how ability develops in each area at both the group and individual levels (Casillas, 2016; Gass, 1984; Sheldon & Strange, 1982; Zampini, 1998). In Casillas (2016), L1 English learners of L2 Spanish labeled a /ba/-/pa/ VOT continuum and completed production tasks weekly over an eight-week immersion program, yielding eight data points. Both perception and production improved significantly over time. However, whereas perception (i.e., perceptual VOT boundary) began to improve from the second session, production, which was operationalized as a VOT boundary by averaging VOT values for Spanish /b/ and /p/ targets, began to improve from the third, which points to a time-lagged relationship.

When interpreting these results, it is important to consider how methodological choices shape findings. For example, it is well documented that task characteristics such as the length of the interstimulus interval may induce either phonetic or phonological processing (Werker & Logan, 1985), and that even when listeners are capable of discriminating non-native L2 contrasts on a phonological processing task, they may not encode those same contrasts at the lexical level (Curtin, Goad, & Pater, 1998). Target structure also plays an important role since stronger transfer from perception training to production has been observed for obstruents than for sonorants and vowels (Sakai & Moorman, 2017), which may be due to the fact that obstruents represent a more concrete articulatory target vis-à-vis manner and place of articulation. To that point, Hao and de Jong (2016) contended that production may exceed perception for L2 sounds that share a close counterpart in the L1. They found that L1 Korean speakers' production of L2 English stops outpaced their perception, but the opposite was true of participants' perception and production of English fricatives (i.e., perception > production).

To summarize, accumulated findings indicate that perception and production are linked, though the strength of the relationship from a statistical perspective depends on an array of

factors, both methodological and linguistic. Although reports have emerged suggesting that changes in perception may precede changes in production (Casillas, 2016) or that perception and production may be asymptotically related in that the latter improves only once the former reaches a certain accuracy threshold (Hanulíková et al., 2012), there has been little research into the temporal dynamics of the perception-production link. At the same time, there is growing awareness that individual differences may lead to distinct perception-production outcomes, and so researchers have become increasingly interested in evaluating group and individual performance.

Target Structure: Stop Consonant Voice Onset Time

Although there are numerous cues to stop consonant voicing in English and Spanish, including f_0 , formant transitions, and closure duration (Benkí, 2005; Zampini & Green, 2001), VOT is one of the most robust (e.g., Lisker & Abramson, 1964). In English, voiced stops are predominantly realized with short positive VOT values between 10 and 30 ms and voiceless stops with longer positive values between 60 and 90 ms (e.g., Flege, 1987; Flege & Eefting, 1987). Consequently, English is characterized as contrasting short- and long-lag stops, though speakers occasionally prevoice /b d g/ (Flege, 1982; B. L. Smith, 1978). On the other hand, voiced stops in Spanish are always prevoiced, which is coded as negative VOT, and voiceless stops are short-lag, produced with short positive VOT (Casteñada, 1986; Rosner, López-Bascuas, García-Albea, & Fahey, 2000; Williams, 1977). In Spanish, voiced stops are typically produced with 70 to 100 ms of prevoicing (i.e., -100 to -70 ms VOT) and voiceless stops with 10 to 20 ms VOT. Thus, Spanish presents a prevoiced/short-lag contrast. In perception, English speakers therefore need to associate short-lag VOT with voiceless stops, shifting their perceptual VOT

boundary to a shorter value to fit the characteristics of Spanish. In production, they need to prevoice voiced stops and reduce VOT in voiceless stops.

The Current Study

The current study is part of a larger longitudinal project on L1 English speakers' perception and production of L2 stops over five data points distributed throughout their second, third, and fourth semesters of college-level Spanish language instruction (Nagle, 2017a, 2017b). This study focuses on the temporal characteristics of the perception-production link in instructed SLA, taking into account three possible configurations: a contemporaneous link, in which perception directly predicts production, delayed models, in which perception predicts production at subsequent data points, and asymptotic models, in which production improves substantially only once perception has reached nativelike accuracy. These relationships are assessed through mixed-effects modeling of the perception and production data, first independently and then using perceptual variables to predict production. Methodologically, this study is among the first to leverage more powerful statistical techniques to understand longitudinal data in L2 phonology. Theoretically, this study opens a window into the core premise of current models of L2 phonetic learning by investigating how perception and production are related over time. The following research questions are addressed:

1. How does learners' perception of L2 Spanish stops develop over two semesters of classroom language instruction?
2. How does learners' production of L2 Spanish stops develop over two semesters of classroom language instruction?
3. What type of model best captures the relationship between learners' perception and production of L2 stops over time?

Method

Participants

Twenty-six L1 English speakers who were enrolled in a second-semester, college-level Spanish course at the time of recruitment participated. On average, participants had begun learning Spanish when they were 14 years old ($M = 14.38$, $SD = 4.11$) and had studied the language for three years in high school ($M = 3.35$, $SD = 3.17$). Three individuals reported experience with another foreign language. Participant 13 (P13) had studied Japanese and had lived in Japan for five years, P15 had studied Italian for one year, and P24 had studied French for four years and had participated in a two-week immersion program. Data from P13 were retained since Japanese presents a short/long-lag contrast more similar to English than Spanish. Data from P15 were likewise retained given his limited contact with Italian. However, P24 was excluded due to her extensive study of French. Five participants withdrew from the study after the first semester of data collection (i.e., after two sessions), reducing the sample size to 20 learners. Six native speakers of Spanish participated once to provide baseline data.

Materials

A cartoon character minimal pair in which /b/ and /p/ occurred in word-initial position and in a stressed syllable was created for the study. In addition to memorizing the two characters, participants worked with a set of 25 basic vocabulary terms, ten verbs and 15 nouns, appearing in their introductory textbook. Learners completed a vocabulary module to ensure that they were familiar with all items relevant to the study. First, they received printed images of the characters and had five minutes to memorize them. After examining the characters, learners advanced to a computerized portion on which they saw an image corresponding to each vocabulary item with the word printed below and heard an audio recording of the word pronounced by a native

Spanish speaker. The program cycled through the terms automatically at a rate of one word every five seconds. At the conclusion of the training, participants took a quiz, selecting the word that correctly described the associated image from among four options. Once learners achieved a perfect score on the quiz, generally after the first attempt, they began the experimental tasks.

Participants completed a reading task and an identification task, in that order, to provide data related to their production and perception of L2 Spanish stops. Both measures were delivered using SuperLab 4.0 software. The reading task was self-paced. Participants saw one sentence per screen, read it aloud, and subsequently advanced to the next trial by pressing any key. There were ten sentences per character (10 tokens for word-initial /b/ and 10 for word-initial /p/) or 20 total. These sentences were similar, but not identical, to the sentences that learners were exposed to on the identification task.

Sentences for the identification task were recorded by a male native speaker of Argentinian Spanish and conformed to one of two patterns: subject-verb-direct object (e.g., *Pafo toca la flauta*, ‘Pafo plays the flute’) or subject-verb-locative phrase (e.g., *Bafo corre en la playa*, ‘Bafo runs on the beach’). The average duration of the recorded sentences was 1.59 seconds ($SD = .07$, range: 1.47 to 1.69), and average VOT was -72 ms for *Bafo* ($SD = 10.99$, range: -89 to -55) and 12 ms for *Pafo* ($SD = .85$, range: 10 to 13). All recordings were normalized to a comfortable volume. On each trial of the identification task, participants heard a tone followed by one of the 20 Spanish sentences and had to select the corresponding image set from four choices. Image sets presented on each response screen differed with respect to the target character and the distractor/control image, which was related to the verb, direct object, or locative phrase, such that participants had to pay attention to the entire sentence to respond correctly. For example, for the sentence *Pafo canta en la cocina* (‘Pafo sings in the kitchen’),

participants might receive image sets depicting *Pafo* (correct) and *Bafo* (incorrect) singing in the kitchen (correct) and in the car (incorrect), or image sets depicting the two characters singing (correct) and eating (incorrect) in the kitchen. Of the 20 trials, half (10 trials) contained verb distractors (i.e., trials on which participants had to select the correct verb in addition to the character) and the other half either direct object (5 trials) or locative phrase (5 trials) distractors. The location of the correct response was counterbalanced by quadrant and character, appearing in the upper left, upper right, lower right, and lower left quadrants on five trials each. Because there were ten trials per character and four quadrants, it was not possible for the correct response to appear in each quadrant an equal number of times per character. Therefore, for a given target character, the correct response appeared in each quadrant at least twice and in two of the four quadrants three times. Following this balancing procedure, two sets of response screens were created and alternated over the course of the study; participants completed the identification task with response set A at the first session, response set B at the second session, A at the third session, and so on.

In addition to the L2 measures, learners recorded a reading task in their L1, English. On this task, participants saw one sentence per screen. Sentences contained either “bark” or “park” in absolute initial position to match the phonetic and sentential context of the Spanish target words as closely as possible. There were ten sentences per word or 20 total. Bark productions were analyzed to determine the extent to which each speaker prevoiced /b/ in English, which was operationalized as the proportion of prevoiced tokens produced out of ten realizations. L1 prevoicing was integrated into analyses to control for the fact that L1 prevoicers tended to produce more prevoiced variants and a greater amount of prevoicing in the L2 than non-prevoicers (for more information, Nagle, 2017a). In other words, this factor was included to

account for the influence of L1 production strategies on learners' L2 production, which could have obscured findings related to the perceptual predictors of interest.

Lastly, learners completed background and language use questionnaires to obtain information related to their language learning experiences and any activities they engage in that were related to Spanish. At the first session they completed a comprehensive language background questionnaire adapted from the Language Contact Profile (Freed, Dewey, Segalowitz, & Halter, 2004). At the third and fifth sessions they completed an abbreviated version focusing on language use outside of their Spanish courses.

Procedure

Participants were recruited from a second-semester Spanish course at a private, midsized university in the United States. Learners participated in five sessions over the course of a calendar year. The first session took place at the midterm of participants' second semester course with sessions occurring at half semester intervals thereafter. Thus, data collection was distributed throughout learners' second, third, and fourth semesters of college-level communicative Spanish language instruction. Both learners and their instructors were asked to provide information on the content of the Spanish courses, particularly as related to pronunciation. Learners reported that they did not receive targeted pronunciation instruction as part of their language courses, and instructors indicated that they did not include pronunciation as part of their daily lessons.

At each session, participants completed the training module, the reading task, the identification task, and the background questionnaire. Completing the reading production measure before the identification perception measure ensured that the aural input that participants received on the identification task did not temporarily enhance their production ability. On the reading task, participants were instructed to speak naturally but clearly.

Recordings took place in a sound-attenuated booth using a dynamic head-mounted microphone (Shure SM10A) connected to a laptop computer via an XLR-to-USB signal adapter (Shure X2u). All recordings were made in Audacity, sampled at 44100 Hz, and exported as .wav files.

On the identification task, participants were instructed to respond as quickly and accurately as possible by left clicking on the box containing the correct image set (i.e., the image set corresponding to the sentence presented aloud). They received sample response screens to familiarize them with what they would see during the task. These screens explained that on each trial they would need to respond based on either (a) the character and the verb, (b) the character and the direct object, or (c) the character and location of the event. Each trial began with a tone signaling the upcoming Spanish sentence, and once participants registered a response, the program automatically advanced to the next trial. Before the experimental phase, participants completed a practice session consisting of four trials drawn randomly from the stimulus set. Trials were randomized such that each participant received a unique order of sentences at each data session.

Perceptual sensitivity as d'

Participants' performance on the distractor portion of the sentence was considered before examining their identification accuracy. Participants had to correctly identify the distractor on at least 80% of trials in order to be included in analyses. All participants performed near ceiling (accuracy > 90%), which indicates that they based their decisions on the entire sentence and not just on the target characters.

Raw perception data were converted to d' , a measure of contrast sensitivity that takes into account response bias. Spanish voiced and voiceless stops would be interpreted as voiced in English based on the VOT values with which they are produced. Consequently, English speakers

must learn to discriminate prevoiced and short-lag stops (i.e., to associate short-lag VOT with phonological voicelessness) to perceive Spanish stops accurately. d' was therefore framed in terms of participants' ability to associate short-lag VOT with L2 /p/. Hit rate was calculated as the frequency with which participants selected *Pafo* when the target was *Pafo* and false alarm rate as the frequency with which participants selected *Pafo* when the target was *Bafo*. Following Macmillan and Creelman (2005), extreme scores were adjusted by a factor of .05 such that a hit rate of 1 was reduced to .95 and a false alarm rate of 0 increased to .05. d' was computed by subtracting the z score of the false alarm rate from the z score of the hit rate: $d' = z(H) - z(F)$. Based on the ten trials per character included in the present study, a maximum d' score of 3.29 was possible for each participant at each session.

VOT Production

Using Praat acoustic analysis software (Boersma & Weenik, 2015), VOT was labeled as the interval between the release burst of the stop and the onset of vocal fold vibration, as indicated by vertical striations in the spectrogram and periodicity in the waveform. In cases where voicing began before the release burst (i.e., during closure), VOT was coded as a negative value to signal prevoicing. A script was used to extract VOT measurements once all files had been labeled.

Results

Modeling L2 stop consonant perception and production

Mixed-effects models were fit using the lme4 package (Bates, Maechler, Bolker, & Walker, 2014) of R (R Core Team, 2016). Data sets were first inspected for normality and outliers using Q-Q plots. In cases where multiple models were fit, model building was cumulative such that each model built upon its predecessor. The more complex model was

accepted only if it improved fit, which was assessed by performing a Chi-squared test on the deviance statistic of nested models. Fit was further evaluated by plotting residuals against fitted values. *P* values were determined using the lmerTest package, which employs the Saitterthwait approximation to compute degrees of freedom for the *t* statistic of fixed effects. Ninety-five percent confidence intervals were also calculated for model parameters, which were only accepted as statistically significant if the 95% CI did not cross zero. Models are presented following reporting procedures outlined in Linck and Cunnings (2015) and Cunnings and Finlayson (2015). Following Flege, Yeni-Komshian, and Liu (1999), nativelike performance was defined as scores falling within two standard deviations of the mean for the native speaker group on the relevant measure.

Models were first fit to participants' *d'* perception scores and VOT production to examine development over time in each domain. Four models were fit to the perception and production data: a null model, a linear growth model treating linear session as a fixed and random effect, and two quadratic growth models treating quadratic session first as a fixed effect and then as both a fixed and a random effect. Separate models were fit to the L2 /b/ and /p/ production data.

The mean *d'* score for the six L1 Spanish speakers was 3.19 (*SD* = .18). Five participants demonstrated nativelike perceptual ability, achieving *d'* scores that fell within two standard deviations of the L1 mean ($d' > 2.84$), over all five data points. Since these individuals had reached perceptual asymptote, their data was excluded from modeling, reducing the sample size to 15 participants who completed all five sessions. Including a quadratic slope as a fixed effect improved fit over the model containing fixed and random effects for linear slopes ($\chi^2(1) = 20.94$, $p < .001$). However, the corresponding by-subject random effect for the quadratic slope did not

improve fit ($\chi^2(3) = 3.68, p = .30$). Therefore, the final model included fixed effects for linear and quadratic slopes and by-subject random intercepts and linear slopes. As summarized in Table 1, participants began the study with an average d' score of 1.23 and improved at a rate of 2.25 units per semester of Spanish. The statistically significant negative coefficient for the quadratic slope indicates that development slowed over time by a factor of -0.80 units per semester of Spanish. Consequently, as displayed in Figure 1, in which the model-estimated trajectory is plotted as a thick black line, most participants approached nativelike perceptual performance by the third or fourth session (i.e., 1 or 1.5 semesters of Spanish).

Table 1. Mixed-effects model of d' development.

<i>Parameters</i>	<i>Fixed effects</i>					<i>Random effects</i>
	Estimate	<i>SE</i>	<i>t</i>	CI 95%	<i>p</i>	By Subject <i>SD</i>
Intercept	1.23	0.19	6.38	0.83, 1.62	< .001	0.51
Linear slope	2.25	0.33	6.84	1.58, 2.90	< .001	0.14
Quadratic slope	- 0.80	0.16	- 5.15	- 1.11, - 0.49	< .001	

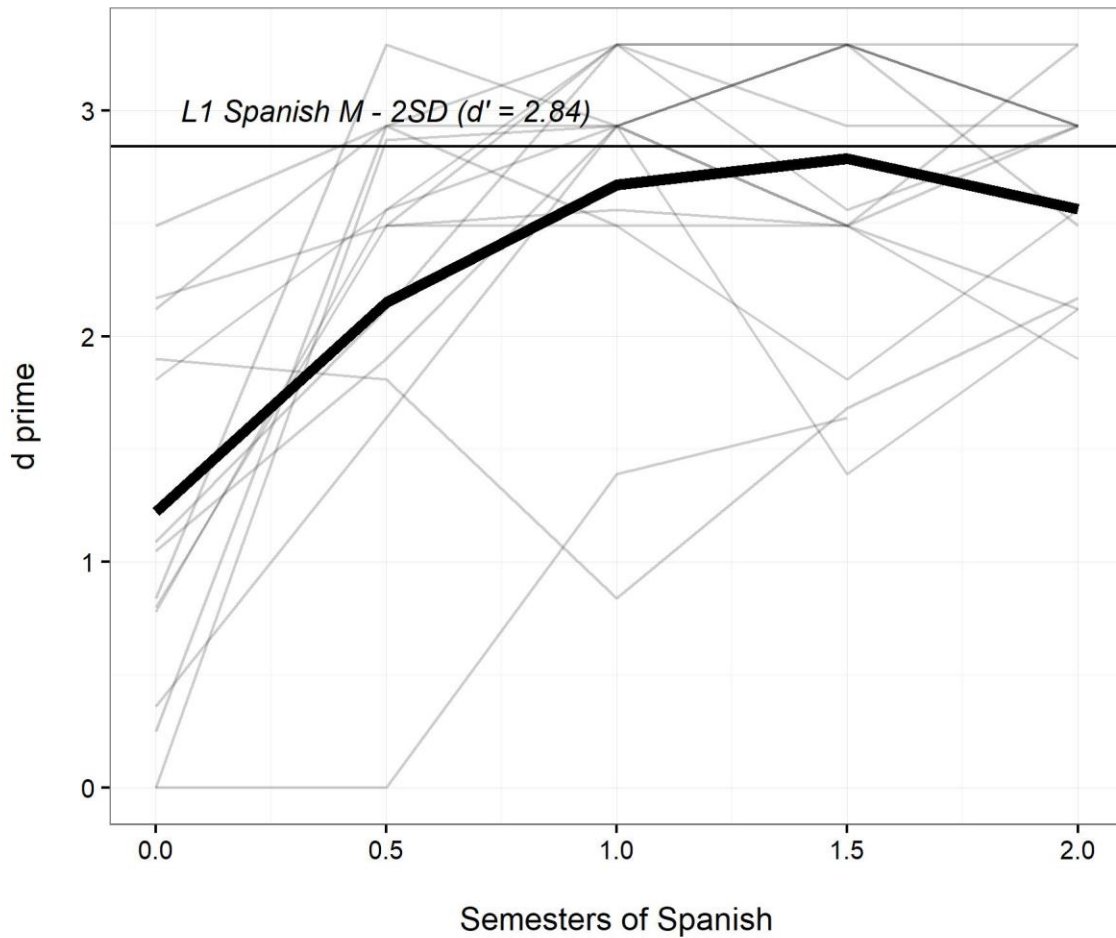


Figure 1. Group (bold line) and individual d' trajectories over two semesters of Spanish. The solid black line at $d' = 2.84$ represents the lower limit of nativelike performance, computed as two SD below the mean of the L1 Spanish group.

A similar approach was followed in modeling VOT production in L2 Spanish /b/ and /p/. For VOT in L2 /b/, there were 18 cases in which VOT could not be accurately measured because the participant coughed or bumped the microphone during recording. Eleven outliers were also identified and removed using Tukey fences, which involved excluding VOT values that fell outside $1.5 \times$ the interquartile range for the group. Consequently, 29 of 1000 observations were excluded from analysis (2.9% of the total data set), resulting in a final data set consisting of 971 observations. Because including a by-subject random effect for the quadratic slope only

marginally improved fit ($\chi^2(3) = 6.87, p = .07$), the simpler model, which included fixed effects for linear and quadratic slopes and random intercepts and linear slopes by subject, was taken as the final unconditional growth model. The frequency with which participants prevoiced /b/ in English, their native language, was integrated into the final model to control for the fact that L1 prevoicers tended to produce more voicing lead in Spanish than participants who predominantly implemented English /b/ as a short-lag stop (Nagle, 2017a). Table 2 summarizes the fixed and random effects structure of the final /b/ model.

At the outset of the study, participants produced L2 /b/ with an average VOT of 10 ms. Whereas the significant linear slope demonstrates that VOT decreased at a rate of -34 ms per semester of Spanish, the positive coefficient for the quadratic slope indicates that the rate of change diminished over time, causing the group trajectory to level off at a VOT value intermediate to those typical for English and Spanish /b/. The significant effect for L1 prevoicing shows that participants who prevoiced more often in English produced greater amounts of prevoicing in Spanish. Figure 2 plots the model-estimated and individual trajectories. As displayed, the group trajectory indicates that although participants clearly began to prevoice Spanish stops, they did not approach a level of prevoicing comparable to the L1 Spanish speakers, whose mean VOT was -94.86 ms ($SD = 22.70$). Nevertheless, some individual learners achieved a mean production that was nativelike.

Table 2. Mixed-effects models of VOT development in L2 Spanish /b/ and /p/.

<i>Fixed effects</i>	<i>Random effects</i> By Subject
----------------------	-------------------------------------

<i>Parameters</i>	Estimate	<i>SE</i>	<i>t</i>	CI 95%	<i>p</i>	<i>SD</i>
<i>/b/ model</i>						
Intercept	10.39	6.12	1.70	-1.41, 22.76	.10	21.66
Linear slope	-33.51	7.64	-4.39	-48.45, -18.54	< .001	22.76
Quadratic slope	9.32	2.70	3.45	4.02, 14.62	< .001	
L1 prevoicing	-54.28	13.93	-3.90	-83.11, -25.44	.001	
<i>/p/ model</i>						
Intercept	46.86	4.46	10.52	37.69, 56.02	< .001	19.44
Linear slope	-23.45	5.18	-4.53	-34.08, -12.77	< .001	20.91
Quadratic slope	8.07	2.29	3.52	3.32, 12.77	.002	9.14

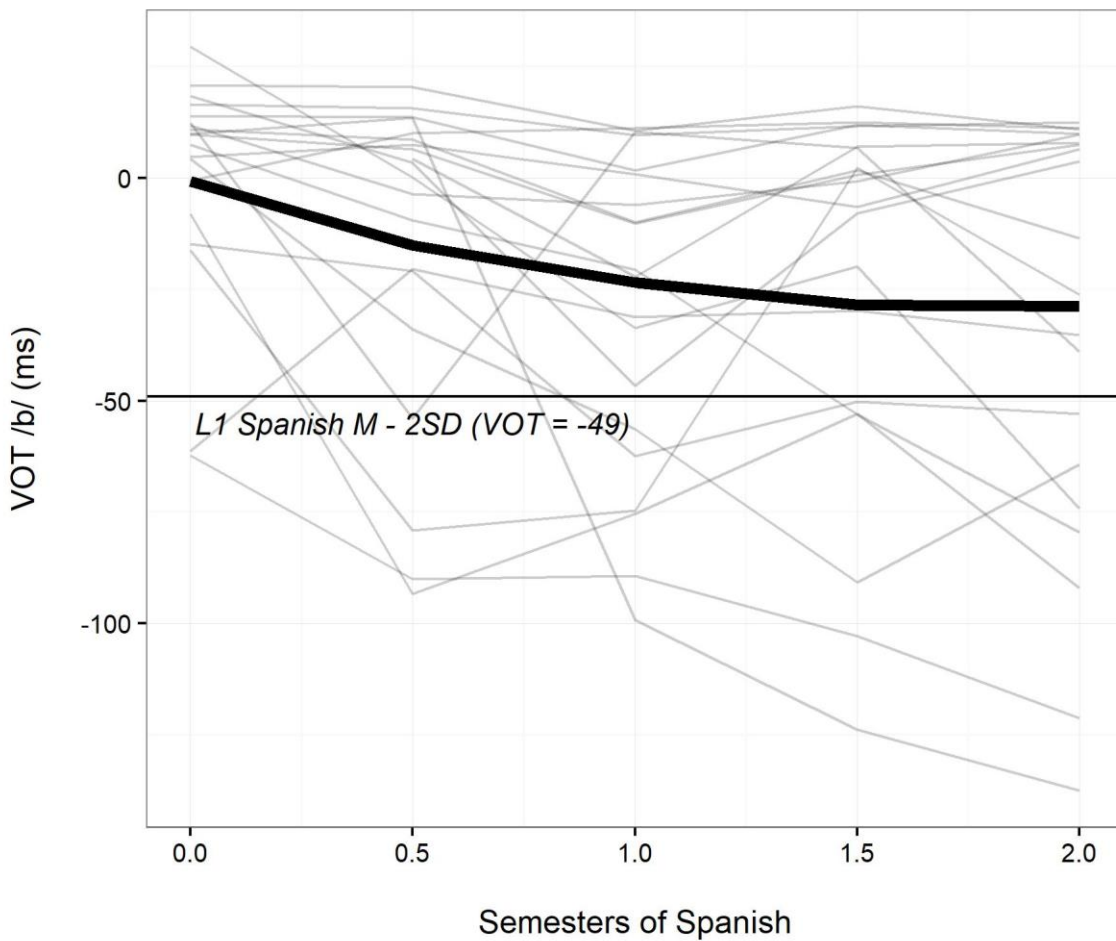


Figure 2. Group (bold line) and individual VOT trajectories in L2 Spanish /b/. The solid black line at VOT = -49 represents the lower limit of nativelike performance, computed as two *SD* below the mean of the L1 Spanish group.

After excluding outliers using Tukey fences ($n = 15$) and cases in which VOT could not be accurately measured ($n = 15$), 970 observations were available for modeling VOT development in L2 /p/. The same taxonomy of models fit to the /b/ data was fit to the /p/ data. In this case, including by-subject random quadratic slopes significantly improved model fit ($\chi^2(3) = 45.80, p < .001$). Table 2 reports the unconditional quadratic growth model for VOT in L2 /p/. The interpretation of the model is similar to the model of L2 /b/. Participants produced Spanish /p/ with an average VOT of 47 ms at the outset of the study. VOT improved significantly over the two-semester period, decreasing at a rate of approximately -23 ms per semester of Spanish. Rate of change was not constant, however, as evidenced by the positive coefficient for the quadratic slope. Consequently, although learners produced shorter, more Spanish-like VOT in L2 /p/, the group trajectory leveled off before reaching the L1 Spanish baseline ($M = 8.51, SD = 1.48$), as Figure 3 illustrates.

To summarize, the perception and production models demonstrate that most development occurred over the first half of the study, as evinced by the significant quadratic slope parameters. With respect to production, group VOT trajectories stabilized between values for Spanish and English stops. Whereas incremental gains were observed in production, perception improved more rapidly with many learners achieving nativelike perceptual ability by the third session. However, the fact that backsliding was common for many individuals indicates that emerging categories and production routines were fragile.

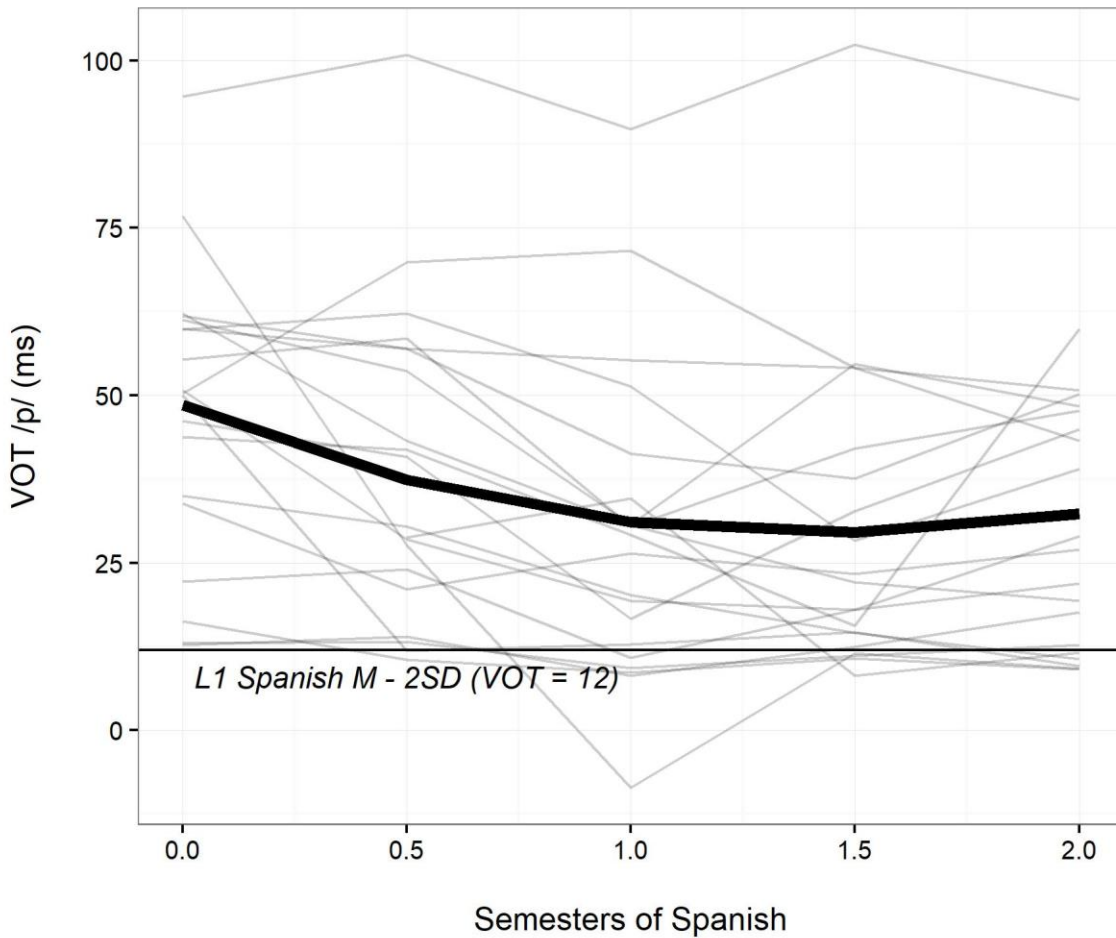


Figure 3. Group (bold line) and individual VOT trajectories in L2 Spanish /p/. The solid black line at $VOT = 12$ represents the lower limit of nativelike performance, computed as two SD below the mean of the L1 Spanish group.

Time-lagged change models

The link between L2 speech perception and L2 speech production is likely to be complex, asynchronous, and dynamic. If a learner's ability to perceive a given contrast improves at time x , it could be that her ability to produce the target phones improves at a later stage, such as at time $x + 1$ or $x + 2$. A series of time-lagged change models was therefore fit to examine perception-production asynchrony, that is, whether gains in perceptual ability predicted contemporaneous or delayed gains in speech production. First, change scores were computed by calculating the

change in each domain from one data point to the next for each participant, resulting in 12 scores per individual, four d' change scores (perception) and four VOT change scores each for L2 /b/ and /p/ (production). Because the goal of these analyses was to determine how changes in perception related to changes in production, participants were removed from the data set once they demonstrated nativelike perceptual ability ($d' > 2.84$). There were five individuals whose perception was nativelike from the first session, six from the second, and 11 by the third. Excluding these cases resulted in a data set consisting of change scores for 15 participants from session 0 to 1, 14 from session 1 to 2, and 9 from sessions 2 to 3 and 3 to 4, or 47 of 80 original observations (20 participants \times four change scores). Perception data was not properly saved for two participants at one of the five sessions, reducing the sample to 45 change scores. Three models were fit to the /b/ and /p/ data separately taking into account three levels of (a)synchrony: a zero-lag model in which changes in perception predicted contemporaneous changes in production, a one-lag model in which changes in perception predicted changes in production at the subsequent data point, and a two-lag model in which changes in perception predicted changes in production two data points later. The lag procedure progressively reduced sample size by excluding one set of change scores per level of lag. Thus, the zero-lag model was based on 45 observations, the one-lag model on 31, and the two-lag model on 17. Due to sample size reduction, a three-lag model was not fit.

All models shared a similar form: for the /b/ models, perception change scores and L1 prevoicing were entered into the model as fixed effects, the latter to control for the relationship between L1 and L2 prevoicing; the /p/ models only contained the perception change scores as fixed effects. When random slopes were fit for the perception change scores, variance approached zero, suggesting that those parameters could be removed. Therefore models were

refit including only the fixed effects for the perception change measures. Table 3 provides an overview of the models with the coefficients for the perception change predictors. The statistically significant negative coefficient for the perception change measure in the one-lag /p/ model (estimate = -7.58, $SE = 2.29$, $p = .003$) provides evidence for a lagged relationship between development in the perception of stop consonant voicing and development of VOT production in L2 /p/. As perceptual ability improved, VOT production in L2 /p/ decreased at the next session, becoming more targetlike, which can be summarized by the following equation: $\Delta VOT/p/x = -0.90 + \Delta d'_{x-1} \times -7.58$. Figure 4 plots $\Delta d'$ at time x on the x-axis and ΔVOT in L2 /p/ at time $x + 1$ on the y-axis including a linear fit line to illustrate the relationship between changes in the former and delayed changes in the latter. No statistically significant relationships emerged between changes in perception and production for L2 /p/ at the other levels of lag (i.e., zero and two), nor did changes in perception predict changes in production for L2 /b/ at any lag level.

Table 3. Summary of fixed effects perception change predictors included in time-lagged production models.

			Fixed effects			
	Cases	Estimate	SE	t	CI 95%	p
<i>/b/ models</i>						
Zero-lag $\Delta d'$	45	-5.83	5.92	-.99	-17.28, 5.62	.33
One-lag $\Delta d'$	30	2.18	5.54	.39	-8.45, 12.82	.70
Two-lag $\Delta d'$	16	12.49	7.57	1.65	-4.38, 25.09	.13
<i>/p/ models</i>						
Zero-lag $\Delta d'$	45	-1.08	2.61	-.42	-6.18, 4.02	.68
One-lag $\Delta d'$	30	-7.58	2.29	-3.13	-12.05, -3.10	.003
Two-lag $\Delta d'$	16	1.99	3.60	.55	-5.03, 9.02	.59

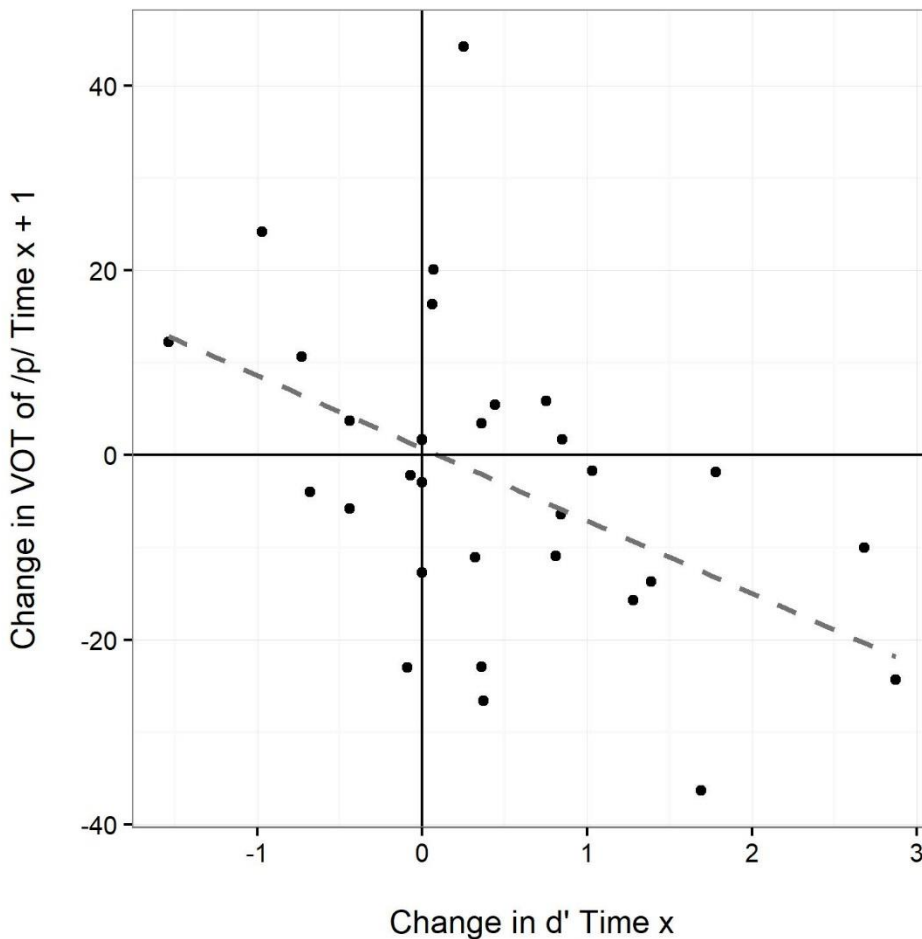


Figure 4. Dot plot of changes in d' (x-axis) and changes in VOT of L2 /p/ (y-axis) at the subsequent session based on 31 change scores. The dotted line represents an estimated linear fit.

Perceptual asymptote models

It could be the case that production improves substantially only once perception has reached nativelike ability. In that case, a shallower developmental trajectory would be expected before perceptual asymptote is achieved. A piecewise linear growth model was fit to the data to evaluate this possibility by recoding linear session into two variables corresponding to data points occurring before and after participants had achieved nativelike perception (i.e., $d' > 2.84$).

The time at which perceptual asymptote was achieved varied among participants: Five participants (25%) were at asymptote at the first data point, nine at the second (45%), 14 at the third and fourth (70%), and 16 at the fifth (80%). Separate models were fit to the L2 /b/ and /p/ data employing the recoded session variables as fixed and random effects. These variables represent pre- and post-asymptote VOT slopes. As in the other /b/ models fit, L1 prevoicing was also included as a fixed effect. As summarized in Table 4, whereas the pre-asymptote slope outpaced the post-asymptote slope for VOT development in L2 /b/ (pre-asymptote estimate = -24.93, $SE = 8.31$, $p = .008$; post-asymptote estimate = -8.60, $SE = 4.88$, $p = .10$), the opposite held true for L2 /p/, although the difference in magnitude between pre (estimate = -5.10, $SE = 2.62$, $p = .07$) and post (estimate = -6.96, $SE = 2.54$, $p = .01$) slopes was small. Consequently, the /b/ model contradicted the perceptual asymptote hypothesis, according to which significant development would be observed only once a high level of perceptual accuracy had been achieved, and the /p/ model provided only weak support given that pre and post slopes differed by a small margin.

Table 4. Summary of models fit with unique slopes pre- and post-perceptual asymptote.

<i>Parameters</i>	<i>Fixed effects</i>					<i>Random effects</i>
	Estimate	<i>SE</i>	<i>t</i>	CI 95%	<i>p</i>	By Subject <i>SD</i>
<i>/b/ model</i>						
Intercept	6.24	6.11	1.02	-5.59, 18.34	.32	22.41
Pre slope	-24.93	8.31	-3.00	-42.62, -8.69	.008	29.91
Post slope	-8.60	4.88	-1.76	-18.38, 1.67	.10	15.77
L1 prevoicing	-49.68	13.97	-3.56	-77.51, -22.11	.002	
<i>/p/ model</i>						
Intercept	42.08	4.49	9.37	33.03, 51.07	< .001	19.71
Pre slope	-5.10	2.62	-1.94	-10.54, .45	.07	9.16

Post slope	-6.96	2.54	-2.74	-11.99, -1.59	.01	9.70
------------	-------	------	-------	---------------	-----	------

Individual differences in the strength and structure of the perception-production link

In evaluating the perception-production link, it is important to examine individual patterns, which not only speak to the range of configurations possible but also reveal whether individual cases support model-estimated group trends. Figures 5 and 6 plot VOT production (y-axis) for /b/ and /p/, respectively, as a function of d' (x-axis) using numbers to represent the five data sessions. In each case, the shaded box represents the L1 Spanish mean \pm 2SD. Drawing upon these figures, Table 5 summarizes change trajectories in d' and VOT for /b/ and /p/ (for by-subject means at each data point, Appendix). As can be observed, in both perception and production, there were learners whose ability was either nativelike at the outset of the study or became nativelike over the course of the study. There were also individual differences in rate of change; for some learners, precipitous shifts occurred, which was more typical in perception than production, whereas for others learning was incremental. Backsliding was common and, at times, resulted in nearly identical initial and final values despite substantial development during intervening data points. For example, for P6, d' decreased from 2.93 at session 2 to 2.49 at session 3 (-0.44) and fell to 1.90 at session 4 (-0.59). In production, P8 produced decreasing VOT in L2 /p/ over the first three sessions. However, despite nativelike VOT at session 2, at session 4 she produced an average VOT of approximately 50 ms, which was comparable to her session 0 production.

The functional form the perception-production link can be assessed at the individual level. A contemporaneous or delayed change model would fit cases in which production improves alongside perception. This type of relationship seems appropriate for P8 for L2 /b/ given that she began to produce substantial prevoicing at session 2 after her perception had

improved significantly from session 0 to 1. A lagged relationship also appears to fit L2 /p/ data for P6, whose perception improved from sessions 0 to 1 and 1 to 2 and production thereafter, from sessions 1 to 2 and 2 to 3. On the other hand, an asymptotic function would be appropriate if production improves once perception has become nativelike. This type of relationship could be applied to P3's production of VOT in L2 /b/ since she discriminated the stops with nativelike accuracy but gradually produced more prevoicing over the study. Likewise, P2, P3, and P9 exhibited nativelike perception and decreasing VOT in L2 /p/. However, there was no learner whose perception reached asymptote at some point during the study and whose production subsequently improved, which would provide the most convincing evidence of an asymptotic relationship. Even though both lagged and asymptotic cases could be tentatively identified in the individual data, many learners displayed no change in production despite nativelike perception (e.g., for L2 /b/: P2, P9 and P16; for L2 /p/, P5), which suggests that a single approach to the perception-production link may be overly simplistic, at least for the sample of learners included in this study.

Table 5. Summary of individual change trajectories in d' and VOT in L2 /b/ and /p/.

id	Perception		Production	
	d'	$d' > 2.84$	VOT /b/	VOT /p/
1	gradual	2	none	abrupt (0-1)
2	nativelike	0	none	gradual
3	nativelike	0	gradual	abrupt (0-1)
4	nativelike	0	none	nativelike
5	gradual	2	none	none
6	gradual	–	none	gradual
7	abrupt (0-1)	–	nativelike	gradual
8	abrupt (0-1)	1	gradual	gradual
9	nativelike	0	none	gradual
10	abrupt (1-2)	–	abrupt (0-1)	abrupt (1-2)
11	abrupt (0-1)	4	abrupt (0-1)	abrupt (0-1)
12	none	–	none	none
13	nativelike	4	abrupt (1-2)	nativelike
14	gradual	2	none	none

15	abrupt (0-1)	2	nativelike	nativelike
16	nativelike	0	none	nativelike
17	abrupt (0-1)	1	abrupt (0-1)	gradual
18	gradual	2	gradual	none
19	abrupt (0-1)	1	abrupt (0-1)	gradual
20	nativelike	1	nativelike	nativelike

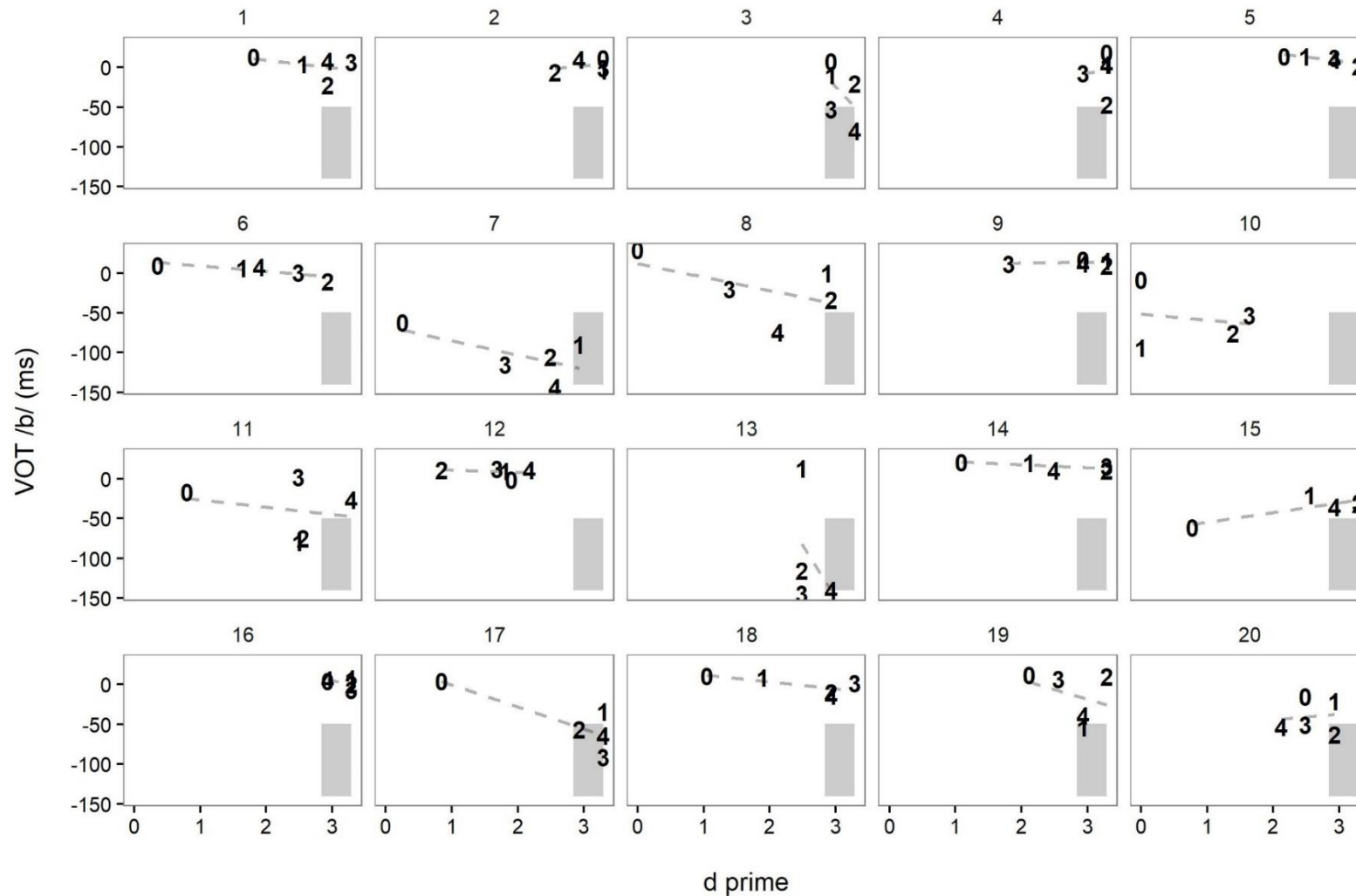


Figure 5. Individual change plots of VOT production in L2 Spanish /b/. Perception is on the x-axis and production on the y-axis. Numbered data points refer to the five data collection sessions (0-4). The shaded box represents nativelike performance. The dotted line is an estimated linear fit line between perception and production scores over time and should be interpreted with caution since the consistency of the apparent relationship must also be taken into consideration.

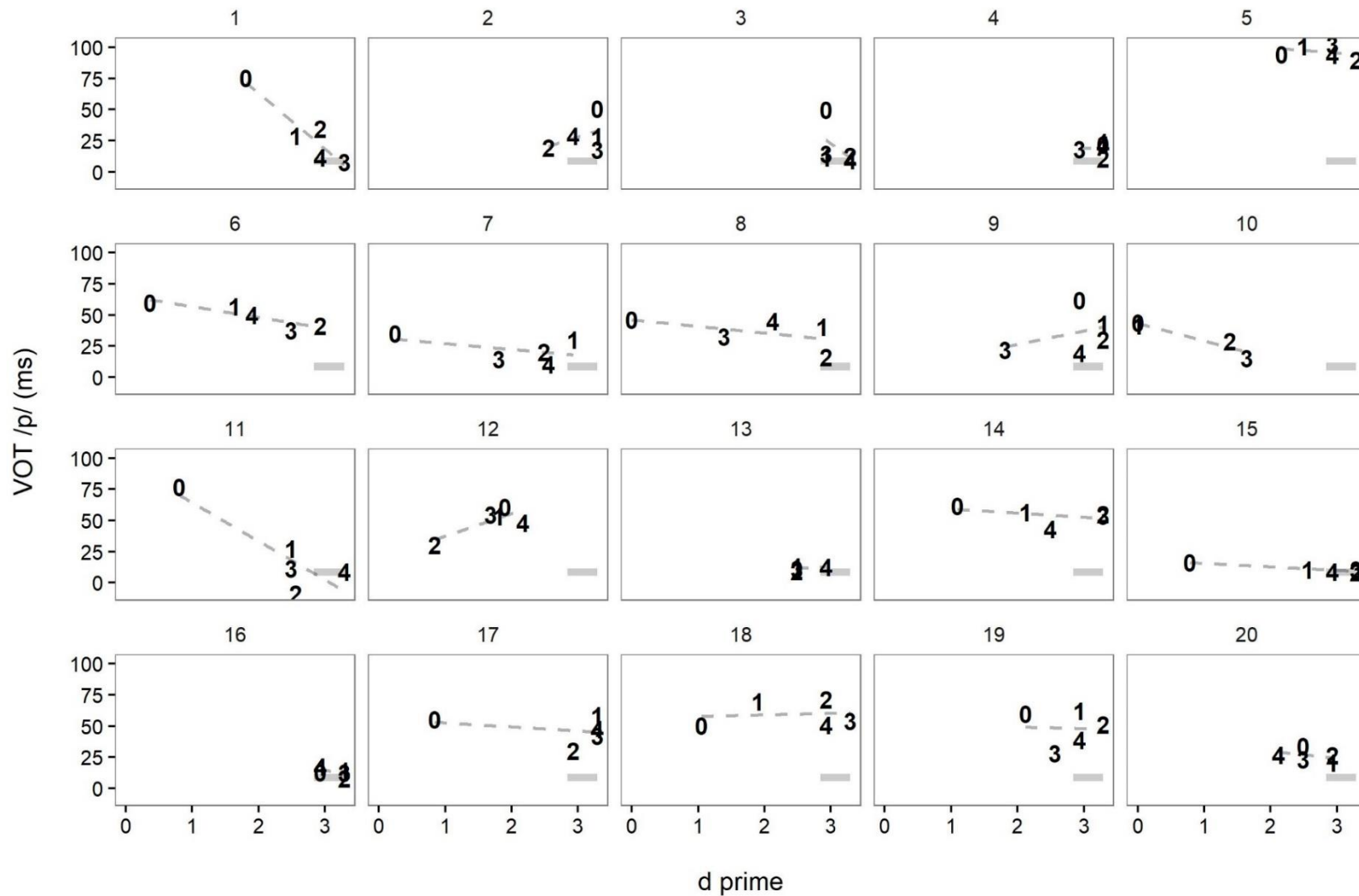


Figure 6. Individual change plots of VOT production in L2 Spanish /p/. Perception is on the x-axis and production on the y-axis. Numbered data points refer to the five data collection sessions (0-4). The shaded box represents nativelike performance. The dotted line is an estimated linear fit line between perception and production scores over time and should be interpreted with caution since the consistency of the apparent relationship must also be taken into consideration.

Discussion

This study examined learners' perception and production of word-initial L2 Spanish stops over five data points distributed throughout their second, third, and fourth semesters of college-level Spanish language instruction, shedding light on the perception-production link. Mixed-effects models fit to d' , a measure of learners' sensitivity to the Spanish prevoiced/short-lag contrast, and VOT in L2 /b/ and /p/ demonstrated significant development in both modalities. Learners' perception of the Spanish voicing contrast improved rapidly, with 70% of participants reaching asymptote (nativelike performance) by the third session. Likewise, the fact that learners produced significantly more prevoicing in L2 /b/, as indexed by increasingly negative VOT values, and shorter VOT in L2 /p/ over the course of the study demonstrates that they were approximating the prevoiced and short-lag VOT ranges typical of native Spanish. However, development began to decelerate before the group trajectory had reached nativelike performance. Moreover, whereas individual trajectories in perception coincided with the model-estimated group trajectory, learners' production of L2 stops was far more variable, which suggests that many individuals struggled to adjust their production despite discriminating Spanish /b/ and /p/ well.

Whereas previous research on the perception-production link has assumed a synchronous relationship, this study examined three types of models: a contemporaneous change model, in which changes in perception predicted changes in production at the same session, time-lagged change models, in which changes in perception predicted delayed changes in production, and an asymptotic model, in which pre- and post-asymptote slopes were compared to determine if significant gains in production were achieved only after participants' performance on the perceptual measure had reached a high level of accuracy. Although the contemporaneous model

relating changes in *d'* to changes in VOT production in L2 /p/ over the same period failed to demonstrate a significant relationship between the two domains, the time-lagged change model suggested that a change in *d'* was significantly related to a decrease in VOT at the following session. In contrast, no significant relationships were observed between changes in perception and contemporaneous or delayed changes in prevoicing of L2 /b/. Exploratory asymptotic models comparing the slope of VOT development before and after learners had achieved nativelike perception did not support the hypothesis that perception must reach a high level of accuracy before production begins to improve. Rather, the pre-asymptote slope for L2 /b/ was significantly greater than the post slope, and the pre- and post-asymptote slopes for L2 /p/ were similar. Individual data provided some support for a range of relationships, including those suggested by the models. However, the diversity of patterns that emerged at the individual level suggests that a single model of the perception-production link may not be possible, or that additional factors may mediate the functional form of the relationship.

According to the SLM, if learners detect differences between an L2 phone and a similar L1 category, then a new L2 category will be formed, facilitating accurate production. Conversely, if the L2 phone is equated with the L1 category, then category formation will be blocked. In this case, the L2 phone will be produced via the L1 category, resulting in an accented production. The SLM could therefore be characterized in terms of both delayed and asymptotic approaches to the perception-production link. On the one hand, increasing accuracy in perception should facilitate increasing accuracy in production but, on the other, nativelike production should only be possible once nativelike perception is observed and possibly sustained over time. The results of the present study partially fit with this perspective since improved discrimination of Spanish stops predicted decreasing VOT in L2 /p/ at the subsequent session. In contrast,

improved perception was not associated with increased prevoicing in L2 /b/ regardless of the type of relationship that was modeled. Consequently, it appears that learners may have begun to form a new category for L2 /p/ but not L2 /b/, which according to the SLM would indicate that they had detected differences between short-lag Spanish /p/ and long-lag English /p/ but not between Spanish and English /b/. There is reason to believe that this may have been the case given the acoustic configuration of the two languages: distinct VOT ranges are observed for voiceless stops but there is considerable overlap with respect to voiced stops since in English prevoiced stops are perceived as voiced, even if they are relatively rare or idiosyncratic. However, this explanation should be interpreted with caution since the perceptual measure employed in the current study considered identification accuracy of both phones simultaneously and is therefore opaque with respect to the perception of individual sounds (i.e., listeners' discrimination of voiced and short-lag stops on the one hand and short- and long-lag stops on the other).

The apparent dissociation between discrimination of the prevoiced/short-lag contrast and production of prevoicing in L2 /b/ could also be related to articulatory factors. Aerodynamic modeling has shown that (pre)voicing is difficult to initiate and maintain in initial stops (Ohala, 1997; Westerbury & Keating, 1986), and L1 acquisition research has demonstrated that children typically produce short-lag stops before acquiring prevoiced (Macken & Barton, 1980) or long-lag categories (Hitchcock & Koenig, 2013), which suggests that a short-lag production is less marked. In fact, when Deuchar and Clark (1996) investigated how a child learning Spanish and English simultaneously acquired voicing contrasts in the two languages, nativelike command of the English system was evident before prevoicing emerged, indicating that prevoiced stops may be the most difficult to acquire. If L2 learning adheres to the developmental stages identified in

L1 acquisition, then learners in the present study might begin to implement prevoicing more consistently at a later stage.

The one session delay between improvements in perception and improvements in VOT production in L2 /p/ can be interpreted within the acoustic/auditory framework of the SLM as the average time required for learners to transduce the acoustic cues available during speech perception to the gestures required for speech production. In theory, through this process, learners were able to adjust the phasing relation between the oral and glottal gestures involved in the production of L2 /p/, resulting in a gradual decrease in VOT over time. However, the lagged results are more easily interpreted within the gestural framework of the PAM-L2, according to which learners directly perceived the gestures needed to produce short-lag stops and were able to begin to implement them in L2 /p/ after a brief period of articulatory hypothesis testing. In either case, acquisition of L2 /p/ was evidently simpler than L2 /b/, which as noted might involve a more substantial delay or require additional articulatory skills that were not assessed in the present design. Yet, in evaluating results with respect to these two models, it becomes clear that neither one directly addresses the time course and mechanisms that operate between perception and production. Thus, it is important to consider what type of model might provide a more comprehensive approach to the perception-production link.

One promising framework is Gambi and Pickering's (2013) Simulation Theory of speech perception. According to Simulation Theory, listeners covertly imitate speech, anticipating the acoustic consequences of upcoming motor commands through a forward model. If the predicted output of the speaker does not match the input the listener receives, then motor commands are adjusted to avoid the likelihood that a similar misperception will occur in the future, which is conceptually analogous to error-driven perceptual learning in the L2LP. Applied to the present

scenario, upon encountering a short delay between stop release and voicing onset in Spanish /p/ (actual acoustic output) rather than the longer delay that would be expected on the basis of English phonetics (predicted output), learners updated their internal system of laryngeal gestures and phasing relations. In this way, they were able to approximate the shorter VOT values typical of native Spanish. Simulation Theory can also account for the lack of relationship between discrimination of prevoiced and short-lag categories and production of L2 /b/: If what learners anticipate is guided by L1 knowledge (or the L1 grammar, as in the L2LP), then both prevoiced and short-lag stops would be anticipated for Spanish /b/ since both occur in English. In this case, predicted (i.e., imitated) and actual outputs would match, and motor commands would not be adjusted. At a more granular level, it could be argued that production is guided by L1 exemplars, which would allow L1 prevoicers to predict a prevoiced realization and non-prevoicers to predict a short-lag realization. L1 prevoicers would then begin to implement L2 stops with prevoicing, exhibiting a more abrupt developmental pattern than non-prevoicers, who would acquire L2 prevoicing more gradually. This approach seems to align with the present data since the shift from short-lag to prevoiced L2 /b/ seemed to be binary for some L1 prevoicers (e.g., P13). Nevertheless, it is unclear how this model would account for the fact that perception and production appeared to be dissociated in some learners in this study.

In fact, previous research indicates that weak or null findings are not uncommon when individual cases are taken into consideration (Baese-Berk & Samuel, 2016; Hanulíková et al., 2012; Kartushina & Frauenfelder, 2014; Kartushina et al., 2015; Peperkamp & Bouchon, 2011; L. Smith, 2001). Scholars have accounted for apparent disruptions or dissociations between perception and production via a range of mechanisms. Baese-Berk and Samuels (2016) found that perception and production may at first compete since producing items during training

seemed to compromise emerging perceptual representations. In terms of a perceptual basis for accented productions, Baker and Trofimovich (2006) have suggested that learners' ability to perceive their own speech and compare it to a model may shape production accuracy. On the other hand, Kartushina and colleagues (2015) have argued for the conditioning role of productive factors such as the articulatory complexity of the target language sound and individual differences in motor control. In fact, all of these factors likely play a role in instructed contexts given that learners must deal with a complex stream of input, in terms of both source (instructor versus peers) and modality (written versus spoken), from the very outset of language study. It is therefore worthwhile to consider what types of factors may promote the development and alignment of perception and production in instructed SLA given that individual differences in metalinguistic strategies, imitation ability, and perceptual aptitude may result in greater perception-production synchrony. While a complete analysis of how these variables may mediate each modality independently as well as their relationship to the perception-production link is beyond the scope of this paper, learners provided a range of details related to their previous language experience, habits, and interests that can be used to evaluate individual profiles from an exploratory descriptive perspective. For example, P15, whose perception improved precipitously and whose production of both phones was nativelike, reported enjoying imitating accents and observed that while she paid attention to her peers' grammar and vocabulary, she focused on her instructors' pronunciation. This may explain her exceptional performance over the study. P16 was intrigued by different accents and like P15 reported imitating her instructors' pronunciation. She also frequently referenced the notion of aptitude, "I believe some people are able to pick up languages really well and adjust their voice/pronunciation to adapt to the accent." Interestingly, despite nativelike performance on the perception task and in terms of her production of L2 /p/,

she doubted her ability in the area: “Although I find the most important aspect [of Spanish] to be speaking, I get flustered. I realize this will never be my strong suit.” Yet, these profiles should not be interpreted in a strictly deterministic sense. To that point, P5 also indicated that she was concerned with her pronunciation (e.g., “The first [thing I care about] is speaking/being intelligible because efforts to learn Spanish are futile if I am not a great speaker,” and “I care a lot about the pronunciation of my words when I speak Spanish”). Unlike P15 and P16, however, her production did not improve despite nativelike perception over the study.

Limitations and Future Research

First and foremost, results must be interpreted with caution due to the small data set employed in the current study, which consisted of 20 learners who completed all five sessions. As other scholars have suggested (Hanulíková et al., 2012), collecting six to eight waves of data from a sample of 100 or more novice language learners would allow for more complex relationships to be modeled and evaluated. In the present study, an asymptotic perception-production link was assessed by fitting separate linear trajectories to data points occurring before and after participants displayed a perceptual asymptote. By increasing the number of sessions, it would be possible to model piecewise and quadratic growth simultaneously, in which case the quadratic function could be constrained to one portion of the model (i.e., to post-asymptote data points). This type of modeling represents an important next step in the area since research has shown that development seems to decelerate over time, either because participants approach ceiling or possibly due to limits on the amount of L2 phonetic accommodation that can take place in the absence of targeted training.

Secondly, because findings for the perception-production link ultimately depend on the measures that are used to assess each modality, care must be taken to select appropriate tasks.

Employing a wider variety of perceptual measures designed to tap into phonetic and phonological representations would clarify how perceived assimilations at both levels shape perceptual development and its relationship to production accuracy. Likewise, including a spontaneous speaking task would be advisable since reading tasks such as the one used in the present study do not tap into the online production routines involved in everyday communicative interactions. From a methodological perspective, equating perception and production is challenging since the former typically involves learners' ability to discriminate L2 contrasts or the location of a perceptual boundary whereas the latter focus on the acoustic characteristics or intelligibility of L2 phones. Thus, it is not always straightforward whether or how perception and production should be compared and to what extent the magnitude of gains in the former should be related to the magnitude of gains in the latter. This concern is partially mitigated by longitudinal research given that it enables a more nuanced discussion of rate and shape of change. Lastly, given that recent research has increasingly highlighted multiple factors that shape performance in each area, such as the compactness with which L1 sounds are produced (Kartushina & Frauenfelder, 2014), it will be important for future studies to consider what variables may mediate the relationship between perception and production. That is, the perception-production link or the degree of synchrony observed between the two domains may depend on individual differences, which means that robust findings would only come to light once those variables are integrated into analyses. Accounting for how individual differences shape the perception-production link will ultimately serve as the basis for revising current models or possibly envisioning new ones altogether.

Conclusion

Via its longitudinal approach, this study examined three possible temporal constraints on the perception-production link, exploring whether discrimination of L2 Spanish stops predicted more accurate VOT production either at the same session (contemporaneous models) or at a subsequent one (time-lagged models), or if perception had to reach a high level of accuracy before significant gains in production were observed (asymptotic models). A significant time-lagged relationship emerged between improved discrimination of the stops and improved VOT production in L2 Spanish /p/ at the subsequent session. Conversely, no significant relationships were found between discrimination and the production of L2 Spanish /b/, which suggests that enhanced perception of the Spanish voicing contrast was insufficient to promote the acquisition of prevoicing in L2 Spanish /b/. This could be due to the fact that initiating and maintaining voicing is aerodynamically challenging in initial stops. Contemporary models of L2 phonetic learning such as the SLM tend to focus on the behavioral outcomes of phonetic category formation and equivalence classification without theorizing in detail about their underlying temporal dynamics. The results of this study suggest that a time-lagged approach to the perception-production link wherein gains in perception transfer to production at a later stage may be appropriate. Although an asymptotic approach would also be possible within current developmental frameworks, production began to improve before perception had reached nativelike levels, contradicting the view that category formation is facilitated through nativelike perception, which in turn promotes production accuracy. However, more research on the functional form of the link involving a variety of target structures and a larger and denser data set is necessary before a more definitive conclusion can be reached.

References

- Baese-Berk, M. M., & Samuel, A. G. (2016). Listeners beware: Speech production may be bad for learning speech sounds. *Journal of Memory and Language*, 89, 23-36.
doi:<https://doi.org/10.1016/j.jml.2015.10.008>
- Baker, W., & Trofimovich, P. (2006). Perceptual paths to accurate production of L2 vowels: The role of individual differences. *IRAL – International Review of Applied Linguistics in Language Teaching*, 44(3). doi:<https://doi.org/10.1515/iral.2006.010>
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). *lme4: Linear mixed-effects models using Eigen and S4*. R package version 1.1.-7. Retrieved from <http://CRAN.R-project.org/package=lme4>
- Benkí, J. R. (2005). Perception of VOT and first formant onset by Spanish and English speakers. In J. Cohen, K. T. McAlister, K. Rolstad, & J. MacSwan (Eds.), *Proceedings of the 4th International Symposium on Bilingualism*. Somerville, MA: Cascadilla Press.
- Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In M. J. Munro & O.-S. Bohn (Eds.), *Second Language Speech Learning: The Role of Language Experience in Speech Perception and Production* (pp. 13-24). Amsterdam: John Benjamins.
- Boersma, P., & Escudero, P. (2008). Learning to perceive a smaller L2 vowel inventory: An Optimality Theory account. In P. Avery, E. Dresher, & K. Rice (Eds.), *Contrast in Phonology: Theory, Perception, Acquisition* (pp. 271-301). Berlin: Mouton de Gruyter.
- Boersma, P., & Weenik, D. (2015). Praat: doing phonetics by computer (Version 5.4.08) [Computer program]. Retrieved from <http://www.praat.org/>

- Borden, G., Gerber, A., & Milsark, G. (1983). Production and perception of the /r/-/l/ contrast in Korean adults learning English. *Language Learning*, 33(4), 499-526.
doi:<https://doi.org/10.1111/j.1467-1770.1983.tb00946.x>
- Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America*, 101(4), 2299-2310.
- Casillas, J. V. (2016). *The longitudinal development of fine phonetic detail in late learners of Spanish*. The University of Arizona.
- Casteñada, V. M. L. (1986). El V.O.T. de las oclusivas sordas y sonoras españolas. *Estudios de fonética experimental*, 2, 92-110.
- Cheour, M., Ceponiene, R., Lehtokoski, A., Luuk, A., Allik, J., Alho, K., & Näätänen, R. (1998). Development of language-specific phoneme representations in the brain. *Nature Neuroscience*, 1(5), 351-353.
- Chládková, K., & Podlipský, V. J. (2011). Native dialect matters: Perceptual assimilation of Dutch vowels by Czech listeners. *J Acoust Soc Am*, 130(4), EL186-192.
doi:<https://doi.org/10.1121/1.3629135>
- Cunnings, I., & Finlayson, I. (2015). Mixed effects modeling and longitudinal data analysis. In L. Plonsky (Ed.), *Advancing Quantitative Methods in Second Language Research* (pp. 159-181). New York: Routledge.
- Curtin, S., Goad, H., & Pater, J. (1998). Phonological transfer and levels of representation: The perceptual acquisition of Thai voice and aspiration by English and French speakers. *Second Language Research*, 14, 289-405.
doi:<https://doi.org/10.1191/026765898674095369>

- Deuchar, M., & Clark, A. (1996). Early bilingual acquisition of the voicing contrast in English and Spanish. *Journal of Phonetics*, 24, 351-365.
doi:<https://doi.org/10.1006/jpho.1996.0019>
- Escudero, P. (2007). Second-language phonology: The role of perception. In M. C. Pennington (Ed.), *Phonology in context* (pp. 109-134). New York, NY.: Palgrave Macmillan.
- Flege, J. E. (1982). Laryngeal timing and phonation onset in utterance-initial English stops. *Journal of Phonetics*, 10, 177-192.
- Flege, J. E. (1987). The production of "new" and "similar" phones in a foreign language: evidence for the effect of equivalence classification. *Journal of Phonetics*, 15, 47-65.
- Flege, J. E. (1995). Second language speech learning: Theory, findings, problems. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research* (pp. 233-277). Timonium, MD: York Press.
- Flege, J. E. (2003). Assessing constraints on second-language segmental production and perception. In N. O. Schiller & A. S. Meyer (Eds.), *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities* (pp. 319-355). Mouton de Gruyter: New York, NY.
- Flege, J. E., Bohn, O.-S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, 25, 437-470.
doi:<https://doi.org/10.1006/jpho.1997.0052>
- Flege, J. E., & Eefting, W. (1987). Production and perception of English stops by native Spanish speakers. *Journal of Phonetics*, 15, 67-83.

- Flege, J. E., MacKay, I. R. A., & Meador, D. (1999). Native Italian speakers' perception and production of English vowels. *Journal of the Acoustical Society of America*, 106(5), 2973-2987. doi:<https://doi.org/10.1121/1.428116>
- Flege, J. E., Yeni-Komshian, G. H., & Liu, S. (1999). Age constraints on second-language acquisition. *Journal of Memory and Language*, 41, 78-104. doi:<https://doi.org/10.1006/jmla.1999.2638>
- Freed, B. F., Dewey, D. P., Segalowitz, N., & Halter, R. (2004). The language contact profile. *Studies in Second Language Acquisition*, 26(2), 349-356. doi:<https://doi.org/10.1017/S027226310426209X>
- Gambi, C., & Pickering, M. J. (2013). Prediction and imitation in speech. *Frontiers in Psychology*, 4. doi:<https://doi.org/10.3389/fpsyg.2013.00340>
- Gass, S. (1984). Development of speech perception and speech production abilities in adult second language learners. *Applied Psycholinguistics*, 5, 51-74. doi:<https://doi.org/10.1017/S0142716400004835>
- Hanulíková, A., Dediu, D., Fang, Z., Bašňáková, J., & Huettig, F. (2012). Individual differences in the acquisition of a complex L2 phonology: A training study. *Language Learning*, 62(S2), 79-109. doi:<https://doi.org/10.1111/j.1467-9922.2012.00707.x>
- Hao, Y.-C., & de Jong, K. (2016). Imitation of second language sounds in relation to L2 perception and production. *Journal of Phonetics*, 54, 151-168. doi:<https://doi.org/10.1016/j.wocn.2015.10.003>
- Hitchcock, E. R., & Koenig, L. L. (2013). The effects of data reduction in determining the schedule of voicing acquisition in young children. *Journal of Speech Language and Hearing Research*, 56(2), 441. doi:[https://doi.org/10.1044/1092-4388\(2012/11-0175\)](https://doi.org/10.1044/1092-4388(2012/11-0175))

- Kartushina, N., & Frauenfelder, U. H. (2014). On the effects of L2 perception and of individual differences in L1 production on L2 pronunciation. *Frontiers in Psychology*, 5.
doi:<https://doi.org/10.3389/fpsyg.2014.01246>
- Kartushina, N., Hervais-Adelman, A., Frauenfelder, U. H., & Golestani, N. (2015). The effect of phonetic production training with visual feedback on the perception and production of foreign speech sounds. *J Acoust Soc Am*, 138(2), 817-832.
doi:<https://doi.org/10.1121/1.4926561>
- Kuhl, P. K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics*, 50(2), 93-107. doi:<https://doi.org/10.3758/BF03212211>
- Kuhl, P. K. (2000). A new view of language acquisition. *PNAS*, 97(22), 11850-11857.
doi:<https://doi.org/10.1073/pnas.97.22.11850>
- Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T. (2008). Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). *Philos Trans R Soc Lond B Biol Sci*, 363(1493), 979-1000. doi:10.1098/rstb.2007.2154
- Kuhl, P. K., Conboy, B. T., Padden, D., Nelson, T., & Pruitt, J. (2005). Early speech perception and later language development: Implications for the "critical period". *Language Learning and Development*, 1(3-4), 237-264.
doi:<http://doi.org/10.1080/15475441.2005.9671948>
- Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12

- months. *Developmental Science*, 9(2), F13-F21. doi:<https://doi.org/10.1111/j.1467-7687.2006.00468.x>
- Linck, J. A., & Cunnings, I. (2015). The utility and application of mixed-effects models in second language research. *Language Learning*, 65(S1), 185-207.
- Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, 20, 384-422.
doi:<http://doi.org/10.1080/00437956.1964.11659830>
- Macken, M. A., & Barton, D. (1980). The acquisition of the voicing contrast in Spanish: A phonetic and phonological study of word-initial stop consonants. *Journal of Child Language*, 7(3), 433-458. doi:<https://doi.org/10.1017/S0305000900002774>
- Macmillan, N. A., & Creelman, C. D. (2005). *Detection Theory: A User's Guide* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
- Mayr, R., & Escudero, P. (2010). Explaining individual variation in L2 perception: Rounded vowels in English learners of German. *Bilingualism: Language and Cognition*, 13(03), 279-297. doi:<https://doi.org/10.1017/s1366728909990022>
- Nagle, C. (2017a). A longitudinal study of voice onset time development in L2 Spanish stops. *Applied Linguistics*. doi:<https://doi.org/10.1093/applin/amx011>
- Nagle, C. (2017b). Individual developmental trajectories in the L2 acquisition of Spanish spirantization. *Journal of Second Language Pronunciation*, 3(2), 219-242.
- Ohala, J. J. (1997). Aerodynamics of phonology *Proceedings of the Seoul International Conference on Linguistics*.
- Peperkamp, S., & Bouchon, C. (2011). The relation between perception and production in L2 phonological processing *Proceedings of the 12th Annual Conference of the International*

- Speech Communication Association (Interspeech 2011)* (pp. 161-164). Rundle Mall: Causal Productions.
- R Core Team. (2016). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org/>
- Rochet, B. L. (1995). Perception and production of second-language speech sounds by adults. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 379-410). Timonium, Maryland: York Press.
- Rosner, B. S., López-Bascuas, L. E., García-Albea, J. E., & Fahey, R. P. (2000). Letter to the Editor: Voice-onset times for Castilian Spanish initial stops. *Journal of Phonetics*, 28, 217-224. doi:<https://doi.org/10.006/jpho.2000.0113>
- Sakai, M., & Moorman, C. (2017). Can perception training improve the production of second language phonemes? A meta-analytic review of twenty-five years of perception training research. *Applied Psycholinguistics*.
- Schertz, J., Cho, T., Lotto, A., & Warner, N. (2015). Individual differences in phonetic cue use in production and perception of a non-native sound contrast. *Journal of Phonetics*, 52, 183-204. doi:<https://doi.org/10.1016/j.wocn.2015.07.003>
- Sheldon, A., & Strange, W. (1982). The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception. *Applied Psycholinguistics*, 3(3), 243-261. doi:<https://doi.org/10.1017/S0142716400001417>
- Smith, B. L. (1978). Effects of place of articulation and vowel environment on 'voiced' stop consonant production. *Glossa*, 12, 163-175.
- Smith, L. (2001). L2 acquisition of English liquids: Evidence for production independent from perception. In X. Bonch-Bruevich, W. J. Crawford, J. Hellermann, C. Higgins, & H.

- Nguyen (Eds.), *The Past, Present, and Future of Second Language Research* (pp. 3-22).
Somerville, MA: Cascadilla Press.
- van Leussen, J.-W., & Escudero, P. (2015). Learning to perceive and recognize a second language: the L2LP model revised. *Frontiers in Psychology*, 6.
doi:<https://doi.org/10.3389/fpsyg.2015.01000>
- Vasiliev, P., & Escudero, P. (2014). Speech perception in second language Spanish. In K. L. Geeslin (Ed.), *The Handbook of Spanish Second Language Acquisition* (pp. 130-145).
Malden, MA: Wiley.
- Werker, J. F., & Logan, J. S. (1985). Cross-language evidence for three factors in speech perception. *Perception & Psychophysics*, 37, 278-286.
- Westerbury, J. R., & Keating, P. (1986). On the naturalness of stop consonant voicing. *Journal of Linguistics*, 22(1), 145-166. doi:<https://doi.org/10.1017/S0022226700010598>
- Williams, L. (1977). The voicing contrast in Spanish. *Journal of Phonetics*, 5, 169-184.
- Williams, L. (1980). Phonetic variation as a function of second-language learning. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), *Child Phonology Volume 2: Perception* (pp. 185-215). New York: Academic.
- Zampini, M. (1998). The relationship between the production and perception of L2 Spanish stops. *Texas Papers in Foreign Language Education*, 3(3), 85-100.
- Zampini, M., & Green, K. P. (2001). The voicing contrast in English and Spanish: The relationship between perception and production. In J. Nicol (Ed.), *One Mind, Two Languages* (pp. 23-48). Malden, MA: Wiley-Blackwell.

Appendix. Individual scores for d' and means for VOT at each session.

id	Session 0			Session 1			Session 2			Session 3			Session 4		
	d'	/b/	/p/	d'	/b/	/p/	d'	/b/	/p/	d'	/b/	/p/	d'	/b/	/p/
2	3.29	12	51	3.29	-4	29	2.56	-6	19	3.29	-1	18	2.93	10	29
4	3.29	19	22	3.29	3	24	3.29	-47	11	2.93	-8	18	3.29	4	22
3	2.93	7	50	2.93	-9	12	3.29	-20	13	2.93	-53	15	3.29	-80	10
9	2.93	17	62	3.29	16	43	3.29	10	30	1.81	12	22	2.93	12	19
16	2.93	5	13	3.29	8	14	3.29	1	8	3.29	-6	12	2.93	6	18
20	2.49	-15	34	2.93	-21	21	2.93	-62	26	2.49	-50	23	2.12	-53	27
5	2.17	14	95	2.49	14	101	3.29	2	90	2.93	12	102	2.93	10	94
19	2.12	12	60	2.93	-54	62	3.29	10	51	2.56	7	28	2.93	-39	39
12	1.90	-1	61	1.81	10	54	0.84	11	31	1.68	13	55	2.17	11	48
1	1.81	14	76	2.56	4	29	2.93	-22	35	3.29	7	8	2.93	8	12
14	1.09	21	62	2.12	21	57	3.29	11	55	3.29	16	54	2.49	11	43
18	1.05	11	50	1.90	9	70	2.93	-10	71	3.29	2	54	2.93	-13	51
17	0.84	4	55	3.29	-34	59	2.93	-56	30	3.29	-91	42	3.29	-64	48
11	0.80	-16	77	2.49	-79	28	2.56	-75	9	2.49	2	11	3.29	-26	9
15	0.78	-61	16	2.56	-20	10	3.29	-31	9	3.29	-30	11	2.93	-35	9
6	0.36	10	60	1.64	7	57	2.93	-10	41	2.49	1	38	1.90	8	50
7	0.25	-62	35	2.93	-90	30	2.49	-105	20	1.81	-115	15	2.56	-143	11
8	.00	29	46	2.87	0	41	2.93	-34	16	1.39	-20	33	2.12	-74	45
10	.00	-8	44	.00	-93	42	1.39	-75	29	1.64	-53	16		-92	60
13		10	13	2.49	14	13	2.49	-115	9	2.49	-144	11	2.93	-140	13
M	1.63	1	49	2.56	-15	40	2.81	-31	30	2.63	-25	29	2.78	-34	33
SD	1.14	24	22	.79	36	23	.66	39	22	.66	46	23	.43	51	22

Notes. Participants are organized from highest to lowest d' score at each session. Shaded cells indicate values that fell within two standard deviations of the L1 Spanish baseline. “M” and “SD” refer to the grand means and standard deviations for the group by session.